

ABSTRACT OF THE DISCLOSURE

A method and system for reducing server load. Requests are received for a service at a first server from a plurality of client devices. The first server decides to identify one or more other servers to provide the service to some of the plurality of client devices. An address is requested of one or more second servers from a server address management entity. A resource identifier is created at the second servers. Some of the plurality of client devices are redirected to get the service from the one or more second servers. A distributed server tree is formed. The first server provides the service to the second servers to be then provided to some of the plurality of client devices, therefore, reducing the load on the first server and providing more efficient service to the plurality of client devices. The servers may be Session Initiation Protocol (SIP) servers.